What is claimed is:

1. A method for separating two or more subsets of polypeptides within a set of polypeptides, comprising:

(a) determining a sequence comparison signature for each amino acid sequence in a set of amino acid sequences, wherein said sequence comparison signature comprises pairwise comparison scores for said amino acid sequence compared to each of the other amino acid sequences in said set;

(b) constructing a distance arrangement comprising said sequence comparison signatures related according to the distance between each of said sequence comparison signatures; and

(c) identifying a first and second cluster of sequence comparison signatures in the distance arrangement, wherein said first cluster comprises sequence comparison signatures for polypeptides having a similar protein fold or biological function, said protein fold or function being different compared to a protein fold or function of polypeptides having sequence comparison signatures in said second cluster.

2. The method of claim 1, wherein said pairwise comparison score is determined by an algorithm selected from the group consisting of Smith-Waterman, BLAST, FASTA, Needleman-Wunsch, Seller or PSI-BLAST.

3. The method of claim 1, wherein said distance comprises a distance selected from the group consisting of a Euclidian distance, exclusive OR distance and Tanimoto coefficient.

4. The method of claim 1, wherein said distance comprises the distance between a sequence comparison signature and a set of sequence comparison signatures.

5. The method of claim 1, wherein said distance comprises a distance selected from the group consisting of a Penrose distance and Mahalanobis distance.

6. The method of claim 1, wherein said cluster of sequence comparison signatures is identified by hierarchical clustering.

7. The method of claim 6, wherein said hierarchical clustering is selected from the group consisting of agglomerative clustering and divisive clustering.

8. The method of claim 1, wherein said cluster of sequence comparison signatures is identified by non-hierarchical clustering.

9. The method of claim 8, wherein said non-hierarchical clustering comprises Jarvis-Patrick clustering.

10. The method of claim 1, wherein said cluster of sequence comparison signatures is identified by cell-based clustering.

11. The method of claim 1, wherein said subset of polypeptides comprises a family of proteins having a common structural fold.

12. The method of claim 1, wherein said subset of polypeptides comprises a family of proteins having a common function.

13. A method for identifying a member of a polypeptide family, comprising:

(a) determining a query sequence comparison signature for an amino acid sequence, wherein said query sequence comparison signature comprises pairwise comparison scores for said amino acid sequence compared to each amino acid sequence in a set;

(b) comparing the distance between said query sequence comparison signature and the sequence comparison signatures for other amino acid sequences in said set, wherein said sequence comparison signatures for other amino acid sequences in said set are clustered into polypeptide families; and

(c) identifying a proximal cluster having one or more sequence comparison signature that has a closer distance to said query sequence comparison signature than the sequence comparison signatures of a distal cluster, thereby identifying the polypeptide having said query sequence comparison signature as being a member of the polypeptide family for the proximal cluster.

14. The method of claim 13, wherein said pairwise comparison score is determined by an algorithm selected from the group consisting of Smith-Waterman, BLAST, FASTA, Needleman-Wunsch, Seller or PSI-BLAST.

15. The method of claim 13, wherein said distance comprises a distance selected from the group consisting of a Euclidian distance, exclusive OR distance and Tanimoto coefficient.

16. The method of claim 13, wherein said distance comprises the distance between a sequence comparison signature and a set of sequence comparison signatures.

17. The method of claim 13, wherein said distance comprises a distance selected from the group consisting of a Penrose distance and Mahalanobis distance.

18. The method of claim 13, wherein said cluster of sequence comparison signatures is identified by hierarchical clustering.

19. The method of claim 18, wherein said hierarchical clustering is selected from the group consisting of agglomerative clustering and divisive clustering.

20. The method of claim 13, wherein said cluster of sequence comparison signatures is identified by non-hierarchical clustering.

21. The method of claim 83, wherein said non-hierarchical clustering comprises Jarvis-Patrick clustering.

22. The method of claim 13, wherein said cluster of sequence comparison signatures is identified by cell-based clustering.

23. The method of claim 13, wherein said polypeptide family comprises polypeptides having a common structural fold.

24. The method of claim 13, wherein said polypeptide family comprises polypeptides having a common function.

25. A method for identifying a polypeptide pharmacofamily, comprising:

(a) determining a sequence comparison signature for each amino acid sequence in a set of amino acid sequences, wherein said sequence comparison signature comprises pairwise comparison scores for said amino acid sequence compared to each of the other amino acid sequences in said set;

(b) constructing a distance arrangement comprising said sequence comparison signatures related according to the distance between each of said sequence comparison signatures; and

(c) identifying separate clusters of sequence comparison signatures in said distance arrangement, wherein said separate clusters comprise sequence comparison signatures for sequences in the same ligand binding family and separate pharmacofamilies.

26. The method of claim 25, wherein said pairwise comparison score is determined by an algorithm selected from the group consisting of Smith-Waterman, BLAST, FASTA, Needleman-Wunsch, Seller or PSI-BLAST.

27. The method of claim 25, wherein said distance comprises a distance selected from the group consisting of a Euclidian distance, exclusive OR distance and Tanimoto coefficient.

28. The method of claim 25, wherein said distance comprises the distance between a sequence comparison signature and a set of sequence comparison signatures.

29. The method of claim 25, wherein said distance comprises a distance selected from the group consisting of a Penrose distance and Mahalanobis distance.

30. The method of claim 25, wherein said cluster of sequence comparison signatures is identified by hierarchical clustering.

31. The method of claim 30, wherein said hierarchical clustering is selected from the group consisting of agglomerative clustering and divisive clustering.

32. The method of claim 25, wherein said cluster of sequence comparison signatures is identified by non-hierarchical clustering.

33. The method of claim 32, wherein said non-hierarchical clustering comprises Jarvis-Patrick clustering.

34. The method of claim 25, wherein said cluster of sequence comparison signatures is identified by cell-based clustering.

35. The method of claim 25, wherein said ligand comprises a nicotinamide adenine dinucleotide-related molecule.

36. The method of claim 35, wherein said nicotinamide adenine dinucleotide-related molecule is selected from the group consisting of oxidized nicotinamide adenine dinucleotide, reduced nicotinamide adenine dinucleotide, oxidized nicotinamide adenine dinucleotide phosphate, reduced nicotinamide adenine dinucleotide phosphate, and a mimetic thereof.

37. A method for identifying a member of a pharmacofamily, comprising:

(a) determining a query sequence comparison signature for an amino acid sequence, wherein said query sequence comparison signature comprises pairwise comparison scores for said amino acid sequence compared to each amino acid sequence in a set;

(b) comparing the distance between said query sequence comparison signature and the sequence comparison signatures for other amino acid sequences in said set, wherein said sequence comparison signatures for other amino acid sequences in said set are clustered into pharmacofamilies; and

(c) identifying a proximal cluster having one or more sequence comparison signature that has a closer distance to said query sequence comparison signature than the sequence comparison signatures of a distal cluster, thereby identifying the sequences having said query sequence comparison signature as being a member of the pharmacofamily for the proximal cluster, wherein the pharmacofamilies for the proximal and distal clusters belong to the same ligand binding family.

38. The method of claim 37, wherein said pairwise comparison score is determined by an algorithm selected from the group consisting of Smith-Waterman, BLAST, FASTA, Needleman-Wunsch, Seller or PSI-BLAST.

39. The method of claim 37, wherein said distance comprises a distance selected from the group consisting of a Euclidian distance, exclusive OR distance and Tanimoto coefficient.

40. The method of claim 37, wherein said distance comprises the distance between a sequence comparison signature and a set of sequence comparison signatures.

41. The method of claim 40, wherein said distance comprises a distance selected from the group consisting of a Penrose distance and Mahalanobis distance.

42. The method of claim 37, wherein said cluster of sequence comparison signatures is identified by hierarchical clustering.

43. The method of claim 42, wherein said hierarchical clustering is selected from the group consisting of agglomerative clustering and divisive clustering.

44. The method of claim 42, wherein said cluster of sequence comparison signatures is identified by non-hierarchical clustering.

45. The method of claim 44, wherein said non-hierarchical clustering comprises Jarvis-Patrick clustering.

46. The method of claim 37, wherein said cluster of sequence comparison signatures is identified by cell-based clustering.

47. The method of claim 37, wherein said ligand comprises a nicotinamide adenine dinucleotide-related molecule.

48. The method of claim 47, wherein said nicotinamide adenine dinucleotide-related molecule is selected from the group consisting of oxidized nicotinamide adenine dinucleotide, reduced nicotinamide adenine dinucleotide, oxidized nicotinamide adenine dinucleotide phosphate, reduced nicotinamide adenine dinucleotide phosphate, and a mimetic thereof.

49. A method for constructing a conformer model, comprising:

(a) determining a sequence comparison signature for each amino acid sequence in a set of amino acid sequences, wherein said sequence comparison signature comprises pairwise comparison scores for said amino acid sequence compared to each of the other amino acid sequences in said set;

(b) constructing a distance arrangement comprising said sequence comparison signatures related according to the distance between each of said sequence comparison signatures;

(c) identifying separate clusters of sequence comparison signatures in said distance arrangement, wherein said separate clusters include sequence comparison signatures for amino acid sequences in the same ligand binding family and separate pharmacofamilies;

(d) determining bound conformations of said ligand bound to the members of a pharmacofamily; and

(e) constructing an average structure of said bound conformations, wherein said average structure is a conformer model of said ligand.

50. The method of claim 49, wherein said pairwise comparison score is determined by an algorithm selected from the group consisting of Smith-Waterman, BLAST, FASTA, Needleman-Wunsch, Seller or PSI-BLAST.

51. The method of claim 49, wherein said distance comprises a distance selected from the group consisting of a Euclidian distance, exclusive OR distance and Tanimoto coefficient.

52. The method of claim 49, wherein said distance comprises the distance between a sequence comparison signature and a set of sequence comparison signatures.

53. The method of claim 52, wherein said distance comprises a distance selected from the group consisting of a Penrose distance and Mahalanobis distance.

54. The method of claim 49, wherein said cluster of sequence comparison signatures is identified by hierarchical clustering.

55. The method of claim 54, wherein said hierarchical clustering is selected from the group consisting of agglomerative clustering and divisive clustering.

56. The method of claim 49, wherein said cluster of sequence comparison signatures is identified by non-hierarchical clustering.

57. The method of claim 56, wherein said non-hierarchical clustering comprises Jarvis-Patrick clustering.

58. The method of claim 49, wherein said cluster of sequence comparison signatures is identified by cell-based clustering.

59. The method of claim 49, wherein said ligand comprises a nicotinamide adenine dinucleotide-related molecule.

60. The method of claim 59, wherein said nicotinamide adenine dinucleotide-related molecule is selected from the group consisting of oxidized nicotinamide adenine dinucleotide, reduced nicotinamide adenine dinucleotide, oxidized nicotinamide adenine dinucleotide phosphate, reduced nicotinamide adenine dinucleotide phosphate, and a mimetic thereof.

61. A method for constructing a pharmacophore model, comprising:

(a) determining a sequence comparison signature for each amino acid sequence in a set of amino acid sequences, wherein said sequence comparison signature comprises pairwise comparison scores for said amino acid sequence compared to each of the other amino acid sequences in said set;

(b) constructing a distance arrangement comprising said sequence comparison signatures related according to the distance between each of said sequence comparison signatures;

(c) identifying separate clusters of sequence comparison signatures in said distance arrangement, wherein said separate clusters comprise sequence comparison signatures for amino acid sequences in the same ligand binding family and separate pharmacofamilies;

(d) comparing the bound conformations of said ligand bound to members of one of said pharmacofamilies;

(e) identifying one or more conformation-dependent properties of said ligand bound to members of one of said pharmacofamilies; and

(f) constructing a pharmacophore model that contains said one or more conformation-dependent properties.

62. The method of claim 61, wherein said pairwise comparison score is determined by an algorithm selected from the group consisting of Smith-Waterman, BLAST, FASTA, Needleman-Wunsch, Seller or PSI-BLAST.

63. The method of claim 61, wherein said distance comprises a distance selected from the group consisting of a Euclidian distance, exclusive OR distance and Tanimoto coefficient.

64. The method of claim 61, wherein said distance comprises the distance between a sequence comparison signature and a set of sequence comparison signatures.

65. The method of claim 64, wherein said distance comprises a distance selected from the group consisting of a Penrose distance and Mahalanobis distance.

66. The method of claim 61, wherein said cluster of sequence comparison signatures is identified by hierarchical clustering.

67. The method of claim 66, wherein said hierarchical clustering is selected from the group consisting of agglomerative clustering and divisive clustering.

68. The method of claim 61, wherein said cluster of sequence comparison signatures is identified by non-hierarchical clustering.

69. The method of claim 68, wherein said non-hierarchical clustering comprises Jarvis-Patrick clustering.

70. The method of claim 61, wherein said cluster of sequence comparison signatures is identified by cell-based clustering.

71. The method of claim 61, wherein said ligand comprises a nicotinamide adenine dinucleotide-related molecule.

72. The method of claim 71, wherein said nicotinamide adenine dinucleotide-related molecule is selected from the group consisting of oxidized nicotinamide adenine dinucleotide, reduced nicotinamide adenine dinucleotide, oxidized nicotinamide adenine dinucleotide phosphate, reduced nicotinamide adenine dinucleotide phosphate, and a mimetic thereof.

73. The method of claim 72, wherein said conformation-dependent property comprises a spectroscopic signal.

74. The method of claim 72, wherein said conformation-dependent property comprises an NMR signal.

75. The method of claim 74, wherein said NMR signal is selected from the group consisting of chemical shift, *J* coupling, dipolar coupling, cross-correlation, nuclear spin relaxation, transferred nuclear Overhauser effect, and any combination thereof.

76. A method for predicting the bound conformation of a ligand bound to polypeptide, comprising:

(a) determining a query sequence comparison signature for an amino acid sequence, wherein said query sequence comparison signature comprises pairwise comparison scores for said amino acid sequence compared to each amino acid sequence in a set;

(b) comparing the distance between said query sequence comparison signature and the sequence comparison signatures for other amino acid sequences in said set, wherein said sequence comparison signatures for other amino acid sequences in said set are clustered into pharmacofamilies;

(c) identifying a proximal cluster having one or more sequence comparison signature that has a closer distance to said query sequence comparison signature than the sequence comparison signatures of a distal cluster, thereby identifying the sequences having said query sequence comparison signature as being a member of the pharmacofamily for the proximal cluster, wherein the pharmacofamilies for the proximal and distal clusters belong to the same ligand binding family; and

(d) obtaining a pharmacophore model of said ligand bound to said pharmacofamily for the proximal cluster, wherein said pharmacophore model comprises a prediction of the bound conformation for said ligand bound to the amino acid sequence having said query sequence comparison signature.

77. The method of claim 76, wherein said pairwise comparison score is determined by an algorithm selected from the group consisting of Smith-Waterman, BLAST, FASTA, Needleman-Wunsch, Seller or PSI-BLAST.

78. The method of claim 76, wherein said distance comprises a distance selected from the group consisting of a Euclidian distance, exclusive OR distance and Tanimoto coefficient.

79. The method of claim 76, wherein said distance comprises the distance between a sequence comparison signature and a set of sequence comparison signatures.

80. The method of claim 79, wherein said distance comprises a distance selected from the group consisting of a Penrose distance and Mahalanobis distance.

81. The method of claim 76, wherein said cluster of sequence comparison signatures is identified by hierarchical clustering.

82. The method of claim 81, wherein said hierarchical clustering is selected from the group consisting of agglomerative clustering and divisive clustering.

83. The method of claim 76, wherein said cluster of sequence comparison signatures is identified by non-hierarchical clustering.

84. The method of claim 83, wherein said non-hierarchical clustering comprises Jarvis-Patrick clustering.

85. The method of claim 76, wherein said cluster of sequence comparison signatures is identified by cell-based clustering.

86. The method of claim 76, wherein said ligand comprises a nicotinamide adenine dinucleotide-related molecule.

87. The method of claim 86, wherein said nicotinamide adenine dinucleotide-related molecule is selected from the group consisting of oxidized nicotinamide adenine dinucleotide, reduced nicotinamide adenine dinucleotide, oxidized nicotinamide adenine dinucleotide phosphate, reduced nicotinamide adenine dinucleotide phosphate, and a mimetic thereof.